THE BEST PATH – FERMAT'S PRINCIPLE

George Szekeres and David Tacon

The diagram on the cover of this magazine illustrates the reflecting property of the parabola: light rays which are parallel to the axis of a parabola are reflected by the parabola in such a way that they pass through a common point. (This point is called the *focus* from the Latin for *fireplace* or *hearth.*) This was known to the ancient Greek geometers and legend has it that Archimedes used parabolic mirrors to set fire to Roman ships by focussing the sun's rays on them during the siege of Syracuse (214–212 BC). This would not have been physically possible but nevertheless the reflecting property of the parabola is of great practical importance and is exploited in the design of many modern instruments, e.g., car headlights, reflecting and radio telescopes, television receiver dishes and so forth. In historical terms the discovery of the reflection property highlights the success of a new program: the mathematization of the study of light. Probably each of us who reads the above statement of the reflecting property and then glances at the cover understands reasonably well what is meant. However a little reflection of the contemplative kind reveals that much is being assumed. The original geometrical result was about straight lines, angles to normals of parabolas and points of intersection etc. We are assuming that this geometric result provides a good interpretation of a physical situation. Some 2,500 years ago the realization that the real world is potentially interpretable through mathematics must have been very dramatic. We wish to discuss one of the most fundamental ideas in the Greek mathematization in the study of light, and trace it through to more modern times.

Euclid (circa 300BC) was undoubtedly not the first geometer to conceive of light as being propagated in straight lines but he was the first to systematically exploit the idea. Euclid asked how the direction of a ray striking the surface of a plane mirror is related to the direction of the reflected ray.

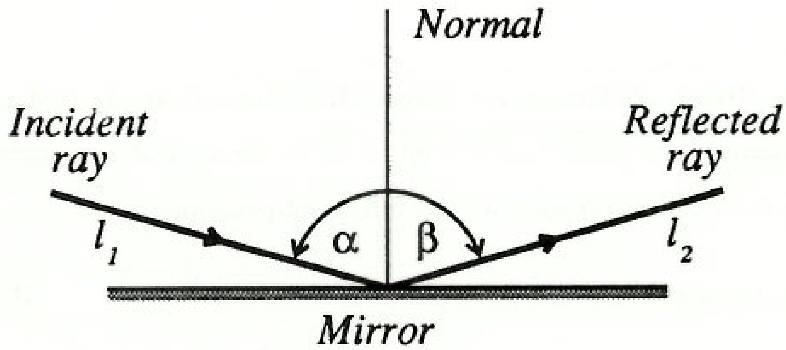

Figure 1

He found experimentally that the reflected ray lies in the plane determined by the incident ray and normal and moreover that the angle of reflection β is equal to the angle of incidence α (see Figure 1). This law is formulated in Euclid's work **Optics** and today is routinely verified in most year 7 science labs. For future reference let's specify Euclid's laws as:

$$\text{Light travels in a straight line} \tag{$E1$}$$

$$\text{Angle of reflection equals angle of incidence} \tag{$E2$}$$

It's worth remembering that in Euclid's day Greek mirrors and measuring instruments would have been very imperfect. How could Euclid have been completely convinced that β was precisely equal to α? He must have held, in the tradition of the Pythagoreans, that Nature is not fortuitous, that her laws have simplicity and elegance. With the courage of conviction he asserted his law held exactly for perfectly plane mirrors.

Not many years after Euclid enunciated these laws Archimedes (circa 287–212 BC) produced a persuasive argument that ($E2$) is true exactly. This argument uses symmetry and is suggested from the observation that if I can see your eyes in a mirror then you can see mine. This means that light transmitted along ℓ_2 is reflected up ℓ_1 (see Figure 1). Once more this can be verified experimentally. In this case the angle of incidence is β and the angle of reflection is α. Archimedes asked: can it be a law that the angle of incidence is (always) larger than the angle of reflection? If this is true then, in the first case $\alpha > \beta$, whilst from the second case, when light is being transmitted in the opposite direction, $\beta > \alpha$. We have a contradiction. Similarly we have a contradiction if the angle of incidence is always smaller. He drew the conclusion that the angles are precisely equal.

Archimedes' use of symmetry is nice but, perhaps surprisingly, he missed seeing a fundamental unifying principle. Heron of Alexandria (circa AD75) is best known today for a formula which bears his name: $A = \sqrt{s(s-a)(s-b)(s-c)}$ is the area of a triangle of sides a, b, c where $s = \frac{1}{2}(a+b+c)$. Heron proposed the following principle:

$$\text{Light takes the shortest path possible} \qquad (H)$$

and proved that both of Euclid's laws (E_1) and (E_2) follow logically from it. This replacement of (E_1) and (E_2) by the single principle (H) is typical of many advancements in science where a single principle replaces a battery of laws of similar or greater complexity. Remember (E_1) implies (H) in certain situations (where there is no reflection involved) so it is attractive to believe (H) in more general situations. Moreover couldn't we conceive of a universe where (E_1) holds yet (E_2) is false?

How did Heron deduce (E_1) and (E_2) from (H)? On the one hand (E_1) follows immediately since the shortest distance between two points A and B in free space is the straight line AB. Now suppose light is to travel from A to B via some point P on a mirror surface.

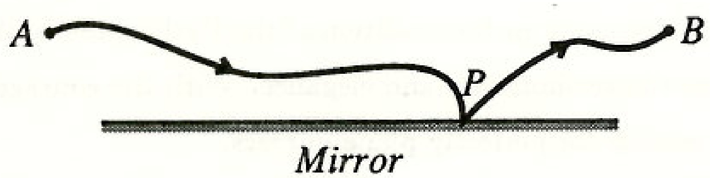

Figure 2

First the lines AP and PB (see Figure 2) cannot be wriggly for the distance AP plus PB cannot be minimal unless both AP and PB are minimal. Therefore we can at least assume the situation shown in Figure 3.

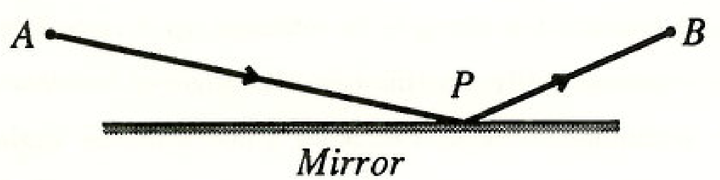

Figure 3

The key to Heron's proof is the introduction of the mirror image B' of the point B. A little elementary geometry reveals that PB is PB' so that $AP + PB$ is minimal when $AP + PB'$ is minimal. But $AP + PB'$ is minimal when APB' is a straight line. Therefore the correct diagram is as in Figure 4.

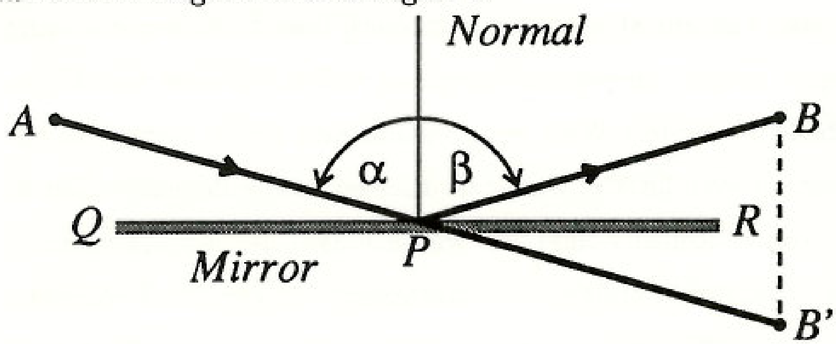

Figure 4

But now $\angle APQ = \angle RPB'$ (vertical opposite angles) and $\angle RPB' = \angle BPR$ (corresponding angles in congruent triangles.) Consequently $\angle APQ = \angle BPR$ and $\alpha = \beta$ as claimed. This subterfuge of Heron whereby he introduced mirror images is often advantageous in solving minimal path length problems and is a favourite trick for competition problem setters! Nevertheless in pure historical terms it is his principle (H) which is the breakthrough.

Not long after the time of Heron the great Alexandrian astronomer Ptolemy (circa 85? – circa 165) found from his observations of the stars that the propagation of light near the earth's surface is not precisely rectilinear. In order to properly understand this bending of light he began to investigate the bending, or refraction, of light when it passes from air to water.

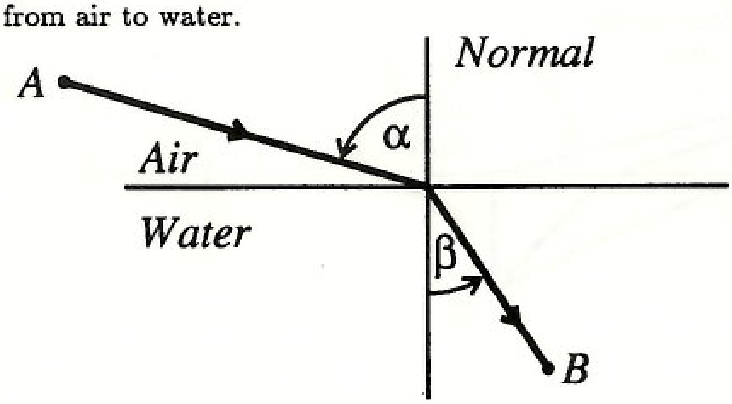

Figure 5

Ptolemy found that the angle of refraction β depends upon α; in mathematical terminology there is a function f such that $\beta = f(\alpha)$. Ptolemy tabulated β for a whole range of α but despite extensive efforts was unable to discover precisely how β depends upon α.

Approximately 1500 years later Ptolemy's incompleted work was taken up by Johann Kepler (1571–1630). Kepler was a genius at sussing out physical laws from overwhelming tables of data. (It took Kepler nearly ten years of incessant toil to discover the 3 laws of planetary motion which bear his name.) With equal enthusiasm Kepler turned to the refraction problem of Ptolemy yet even he was unable to discover how β depends upon α.

One of the livelier questions of Kepler's time was whether light has a finite velocity, or, equivalently, whether the propagation of light is instantaneous. A friend and colleague of Kepler, Galileo Galilei (1564–1642), tried to answer this question by signalling with lanterns between mountain tops. As near as he could tell light was propagated instantaneously. These speculations and experiments captured the attention of Pierre Fermat (1601–1665). Suppose, thought Fermat, that light is propagated with finite velocity. Then the supposition, for free space, that light takes the shortest time is equivalent to Heron's principle that it takes the shortest path. Now suppose, Fermat imagined, that the velocity of light is constant in any given medium but different in different media. In particular suppose that the velocity of light in water is different from its velocity in air. The shortest path is then not the quickest if the ray travels in both air and water. Does it take the quickest? If so in each medium Heron's shortest path principle still holds and refraction is explicable.

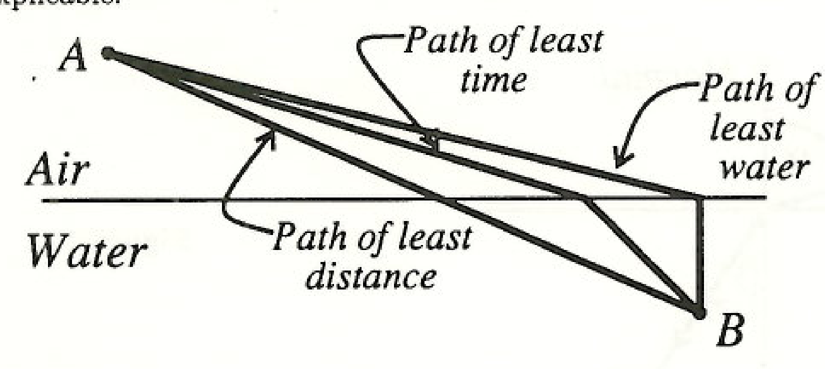

Figure 6

Fermat's imaginative guess, which is known today as Fermat's Principle, postulates:

$$\text{Light follows the path of least time} \qquad (F)$$

What is the relationship between α and β if (F) is true? We will easily work this out using the Calculus but Fermat's original argument was much more subtle and complex. This was simply because the Calculus did not exist in Fermat's day. Fermat himself did much of the groundwork for the development of the Calculus but we really have to wait until the time of Newton and Leibnitz before we have the battery of simple rules of the Calculus that high school students can apply today. If you do not know any Calculus you will have to accept what follows on trust but make sure you come back to the derivation one day. We introduce coordinates as in the diagram of Figure 7.

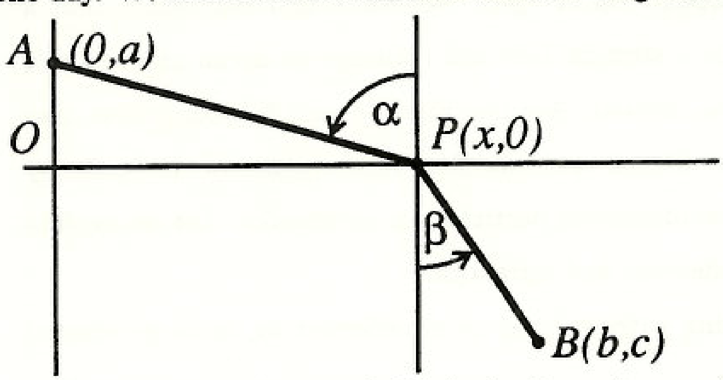

Figure 7

Suppose v_1 is the speed of light in air, v_2 its speed in water. Then the time taken for the light to travel from A to B depends upon x and is given by

$$T(x) = \frac{\sqrt{a^2+x^2}}{v_1} + \frac{\sqrt{(b-x)^2+c^2}}{v_2} \qquad (\text{where } 0 \le x \le b).$$

The derivative is

$$\begin{aligned}\frac{dT}{dx} &= \frac{x}{\sqrt{a^2+x^2}}\frac{1}{v_1} + \frac{-(b-x)}{\sqrt{(b-x)^2+c^2}}\frac{1}{v_2} \\ &= \sin\alpha\frac{1}{v_1} - \sin\beta\frac{1}{v_2}\end{aligned}$$

When $T(x)$ is a minimum we must have $\frac{dT}{dx} = 0$. Therefore if Fermat's principle holds we must have

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}$$

which implies that we should find that $\dfrac{\sin\alpha}{\sin\beta} = k$ where k is a constant. Is this the correct relationship between α and β? The answer is "yes", the experimental data of Ptolemy, and more accurate later data, does satisfy this relationship. Nowadays the law of refraction is usually attributed to Snell and Descartes although they discovered it years after Fermat. Should we give Fermat the Nobel Prize? If we lived in Fermat's time we would probably be very reluctant. After all, for all we know it might be the case that light is propagated instantaneously. Perhaps we should wait for Newton (1642– 1727).

With Newton science came of age. The solar system is a gigantic piece of clockwork, and Newton had discovered how it ticks. Newton's mechanics is the key to everything around the sun, must it not be the key to everything under the Sun? Newton thought that the optics of Euclid, Heron and Fermat could be explained mechanistically. The first thing to explain is the rectilinear propagation of light. Newton's first law states that a body moving with uniform velocity in a straight line will continue to do so unless acted upon by external forces to change that motion. Are not Euclid's and Newton's first laws remarkably similar? Newton made the former an immediate consequence of the latter by postulating that a ray of light consists of minute particles, or *corpuscles*. Let us explain very briefly how Newton explained reflection and refraction.

Suppose that a corpuscle travelling with velocity v_1 is reflected by, or is ricocheted off, a mirror as shown in the diagram

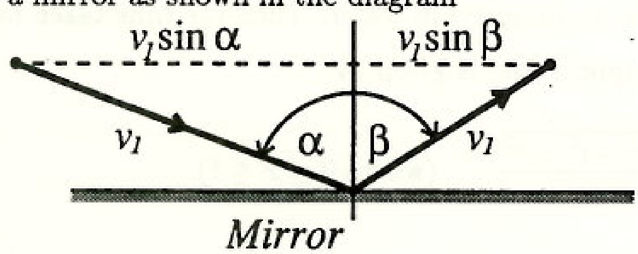

Figure 8

Newton argues that the resistance to penetration of the mirror is solely by forces acting normally outwards. Such forces have no horizontal components and so, by Newton's 1st law, the horizontal component of velocity of the corpuscle is not changed. Equating these components of velocity parallel to the mirror gives

$$v_1 \sin\beta = v_1 \sin\alpha$$

and hence $\beta = \alpha$.

Newton explains refraction in the same way. He continues to assume that the only forces resisting penetration are perpendicular to the surface of the water.

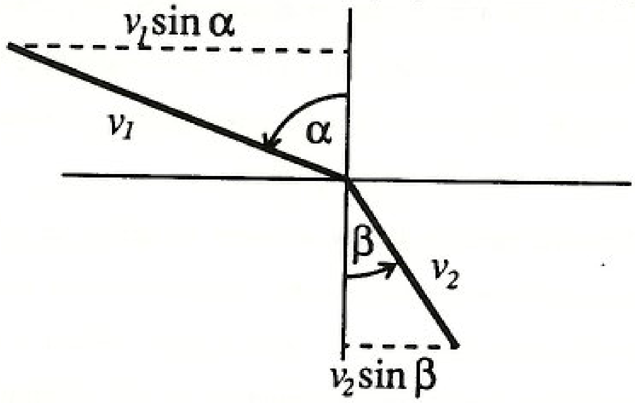

Figure 9

As in the case of reflection the horizontal components of velocity must be the same. Therefore

$$v_2 \sin\beta = v_1 \sin\alpha$$

and $\dfrac{\sin\alpha}{\sin\beta} = \dfrac{v_2}{v_1} = a$ constant, as before.

But it is not quite the same as Fermat's result! If Fermat's constant is k Newton is arguing that the constant should be $1/k$. At least one of the explanations is wrong even though both results successfully explain Ptolemy's data. Now from the experimental data it is apparent that

$$\frac{\sin\alpha}{\sin\beta} > 1.$$

So if Fermat is right light slows down when it is transmitted through water whilst if Newton is right it speeds up. Common sense says a corpuscle should slow down in water but nevertheless Newton was able to think up an argument that it should speed up. This conflict of opinion lasted more than a century until technological advances made it possible to show experimentally that light travels more slowly in water than in air. Nevertheless Newton's argument turns out not to be quite as absurd as it seems: in media such as water or glass there is a second kind of velocity of waves, the so-called *group velocity* – a rather complicated concept, but one that is relevant in Newton-type mechanical arguments – which does behave exactly like the velocity that Newton calculated.

Fermat's principle may strike you as being little more than a lucky guess, made at a time when it was not even known whether light takes any time at all to get from A to

B. Far from it. Imagine you have a complicated optical system made up of media whose composition varies from place to place such as the atmosphere or glasses of varying optical density. Fermat's principle is still valid in all its generality and allows us to calculate the path that light will follow to get from A to B. As such it has acquired great significance and immense use in geometrical optics, the science of optical systems.

How does light "know" at point A which path to follow in order to arrive at a far away point B (which could be the moon or a distant galaxy) in the shortest possible time? A puzzling philosophical question, but of course it is put the wrong way. When a ray of light starts off at A in a certain direction it has "no idea" whether it will ultimately land at B or not, but if it happened to arrive at B, one can verify **after** the event that it has followed the path of shortest possible duration. As such, Fermat's principle doesn't "explain" anything but it certainly supplies the correct answer – provided of course that you are able to handle the difficult calculations.

In more recent times Fermat's principle, in a modified form, has acquired great significance for all mechanics. In the post – Newton era when more refined techniques became necessary to handle the ever more complicated problems thrown up by Newtonian mechanics, principles were formulated which were in much the same spirit as Fermat's principle. The great Irish mathematician Sir William Rowan Hamilton (1805–1865) gave it its final form. Suppose a particle such as an electron (a charged elementary particle), moves in a complicated magnetic field from point A to point B. According to Hamilton's principle among all possible paths from A to B it will follow that one for which the total "action" is the smallest. Action is a quantity in mechanics which is roughly speaking the momentum of the particle (momentum = mass times velocity) at a particular point P of its path multiplied by its distance to a neighbouring point P' of the path. The total action is the sum of all these bits of action, and can be expressed as an integral and hence calculated. The only difference between Fermat's and Hamilton's principles is that "time" is replaced by "action". Hamilton himself was fully aware of the analogy of the two principles. In more recent times this analogy has been carried even further and became one of the cornerstones of quantum mechanics – the mechanics of the atomic world. A far cry indeed from the beginnings laid down by Fermat.

A detailed explanation of all this would of course lead us far beyond our present scope which was merely to give you a feel for the remarkable insights of Fermat. Nowadays he is chiefly remembered for his work in number theory where his "modest" insights have generated in this and the previous century more spectacular advances in modern mathematics then by any other mathematician of the pre-Newton times.

GENTLEMAN, SOLDIER AND MATHEMATICIAN

Descartes (1596-1650)

The good old days. A child philosopher but no prig. Inestimable advantages of lying in bed. Invigorating doubts. Peace in war. Converted by a nightmare. Revelation of analytic geometry. More butchering. Circuses, professional jealousy, swashbuckling, accommodating lady friends. Distaste for hell-fire and respect for the Church. Saved by a brace of cardinals. A Pope brains himself. Twenty years a recluse. The Method, *Betrayed by fame. Doting Elizabeth. What Descartes really thought of her. Conceited Christine. What she did to Descartes. Creative simplicity of his geometry.*

THE PRINCE OF AMATEURS

Fermat (1601-65)

Greatest mathematician of the seventeenth century. Fermat's busy, practical life. Mathematics his hobby. His flick to the calculus. His profound physical principle. Analytic geometry again. Arithmetica and logistica. Fermat's supremacy in arithmetic. An unsolved problem on primes. Why are some theorems 'important'? An intelligence test. 'Infinite descent.' Fermat's unanswered challenge to posterity.

– from the table of contents in **Men of Mathematics** by E.T. Bell.